COMPOSITIONS AND METHODS FOR TARGETED GENE INSERTION

This application claims priority to U.S. Provisional Application No. 60/138,968, Filed June 8, 1999, the entirety of which is incorporated by reference herein.

FIELD OF THE INVENTION

This invention relates to the field of molecular biology and manipulation of the eucaryotic genome. In particular, the invention provides a novel system and DNA constructs for integrating heterologous DNA segments at selected locations in target genomes.

BACKGROUND OF THE INVENTION

Various scientific articles are referred to in parentheses throughout the specification, and complete citations are listed at the end of the specification. These articles are incorporated by reference herein to describe the state of the art to which this invention pertains.

The ability to create a null mutation in a specific gene can provide an unambiguous test of the functional role of its gene product in an organism. Creating a null mutation has obvious advantages over approaches utilizing antisense transcripts since null mutations do not present problems such as incomplete suppression of the target gene product and unknown specificity of its effects. In addition, the dominant nature of the antisense approach for gene suppression is one major drawback. Thus, if the gene of interest is critical for survival or fitness of the organism, one may inadvertently select against transformants that have

effectively "shut down" the expression of the target gene. Alternatively, one may select for spontaneous second site mutations that compensate for the defect caused by the gene suppression. In contrast, the recessive nature of targeted gene insertion via homologous recombination should avoid these concerns. The first targeted progeny should be in the heterozygous state and in most cases, a wild-type phenotype would be expected. The phenotype(s) caused by the loss of the targeted gene can be studied in the homozygous progenies of subsequent generations. In this way, even housekeeping genes that are essential can be studied as embryo lethals.

Another type of reverse genetics approach is the so-called "gene machine" screens, in which a large collection of random T-DNA or transposon integration events are screened by PCR to identify insertions in or near the locus of interest (1, 4, 6). Although this technique has been successful in the identification of insertion mutations for genes of interest, a routine gene targeting approach should be more versatile in the directed mutagenesis of specific genes. For example, approaches such as gene swapping, the so-called "knock-in" mutation, or any other precise alterations at the locus of interest are not possible with the "gene machine" approach.

Although gene targeting has become a well-defined technique in mouse research (2), the specific disruption of a non-selectable locus in higher plants has not been reported until recently. Most of the earlier work on gene targeting in higher plants involved the repair/mutation of a selectable marker gene (8-10). The observed frequencies of recombination using that method were invariably low (9).

For ectopic expression studies in which a

desired gene product is to be produced, the specific targeting of the transgene to a preselected locus should minimize variations in transgene activity due to position effects and/or cosuppression. Since neither one of these phenomena is well understood at the mechanistic level, the currently available strategy to obtain the desired expression levels and pattern is to screen a large number of independent transformed lines. This can be laborious and time-consuming. Insertion of the desired transgene into a preselected locus of the genome would avoid these technical problems and help streamline the process of engineering a desired trait in the plant of interest.

For commercial purposes, specific manipulation of the genome through gene targeting should also greatly facilitate the process of registration for the transgenic materials. At the present time, clinical trials to demonstrate the safety of transgenic crops are needed for any new transgenic plant materials to be released as commercial products. One of the major reasons for this requirement is to safeguard against production of allergens or toxins in new transgenic lines that may result from inadvertent mutations due to random insertion events. Targeted gene insertion would obviate this concern and thus result in substantial savings of time and resources for this phase of product commercialization.

One approach to targeting employed a general gene-targeting construct using a kanamycin resistance gene (*NptII*) as a positive selection marker (7). Polylinkers were designed in the flanking regions of this marker to facilitate the cloning of genomic fragments from the target of interest. To facilitate detection of illegitimate insertion events, a GUS expression unit, which is a screenable marker, was inserted outside of the homologous regions. In the event of a double cross-over

targeting event, the resultant plant cell is kan^{r}GUS$^-$. The feasibility of this vector was demonstrated by targeting of the *NptII* marker into the *TGA3* locus, which encodes a transcription factor of *Arabidopsis* (7). The absence of a negative selection marker enabled a direct estimate of the relative frequency of the targeting event to that of random insertion. In one set of experiments, this number was in the range of 1 to 2 targeting events per 2,580 transformants. This methodology was used to create the first "knock-out" *Arabidopsis* plant (3). In that work, a targeted event at the *AGL5* locus (which also encodes a transcription factor) was recovered in 1 of about 750 transformants (3).

Although successful targeted gene insertion was achieved using the strategy outlined above, that strategy remains limited in several important respects. First, a large number of independent transformation events are needed in order to obtain the infrequent homologous recombination events rather than the more frequent random integration events. This limitation precludes use of the current methodology in most species of agronomic interest. Second, the currently available method has no means for limiting the number of complex integration events that could occur because the number of recombination substrates is greater than one per cell; nor does it allow for application of negative selection strategies which could expedite the process of detecting the rare targeted integration events.

Despite the obvious value of targeting mutations to specific, selected locations within the genome, sometimes it is yet desirable to screen for particular phenotypes among a population of random mutations. The method of activation tagging (14, 15) for creating random mutants has proven valuable. The method involves tagging genes at random by the insertion of DNA

constructs comprising a selection marker gene and transcriptional promoters which are able to transactivate the expression of genes in the vicinity of the insertion. The result is the generation of dominant mutants which survive in the presence of the selection agent and which have been useful in studying genetic influences on plant growth substances, polyamine metabolism, signal transduction by cytokines and abscisic acid, for example. The method retains the disadvantage of requiring large numbers of transformants which precludes application to many agronomically important plants.

SUMMARY OF THE INVENTION

The present invention provides a new process and new DNA constructs and vectors for targeted manipulation of eucaryotic genomes. One key feature of this novel approach for gene targeting is the generation of recombinant substrates through the deployment of transposable elements. In a preferred embodiment, it provides a key advantage by minimizing the number of recombination substrates to one per cell. Among other advantages of the approach, it solves two general problems associated with currently available methods. First, it minimizes the number of independent transformation events required to obtain the infrequent homologous recombination events rather than the more frequent random integration events. This enables the application of gene targeting technologies to more organisms of interest. Second, the novel deployment of powerful negative selection strategies streamlines the recovery of low-frequency homologous recombination events by suppressing or eliminating complex integration processes.

According to one aspect of the invention, a general DNA construct for producing a gene targeting

construct is provided. The DNA construct is bounded by termini within which are a pair of DNA substrates for a selected transposase. These DNA substrates contain between them the following elements: (1) a first cloning site and a second cloning site for insertion of one or more additional DNA segments, wherein the first cloning site and the second cloning site have disposed between them a positive selection gene encoding a gene product that confers to the cells a selectable phenotype comprising resistance to a positive selection agent that is deleterious or lethal to cells having genomes in which the DNA construct has not integrated; and (2) a negative selection gene disposed between one of the DNA substrates for the selected transposase and either the first cloning site or the second cloning site, but not between the first cloning site and the second cloning site, the negative selection gene conferring to the cells a selectable phenotype comprising susceptibility to a negative selection agent, to which cells having genomes in which the DNA construct has not integrated are not susceptible. Optionally, the DNA construct contains a detectable marker gene encoding a detectable gene product. The detectable marker gene is operably inserted in the DNA construct relative to one of the DNA substrates for the selected transposase such that, upon excision of the DNA construct from a genome by the action of the transposase, the detectable gene product is no longer detectable. Preferably, the detectable marker gene is inserted in the construct such that one of the DNA substrates for the selected transposase is located within the gene, between its promoter and coding sequence.

The DNA construct described above can be used for random insertion of a gene of interest into a genome. In accordance with a significant aspect of the invention,

however, it is adapted for integrating a heterologous DNA segment at a pre-determined location of a genome. The adaptation comprises inserting a first targeting segment in the first cloning site and a second targeting segment in the second cloning site. Each targeting segment comprises a DNA sequence substantially homologous to sequences in the genome comprising or flanking the pre-determined location. The presence of the targeting segments enables the DNA construct to integrate into the genome at the pre-determined location by homologous recombination.

In a preferred embodiment, the above-described DNA constructs are operably inserted into a vector for transforming cells. Preferably, for the transformation of plant cells, the vector is an *Agrobacterium* vector.

According to another aspect of the invention, the following method for inserting a heterologous DNA molecule into a pre-determined location on a plant genome is provided, utilizing the above-described *Agrobacterium* vector.

Step 1. Cells are transformed with the vector. The DNA construct can integrate into the genome randomly (more frequent) or by homologous recombination with the targeted genomic DNA sequence (less frequent). At this stage, transformants with random insertions are selected based on their resistance to the positive selection agent and sensitivity to the negative selection agent, and (optionally) their expression of the detectable gene product. In a preferred embodiment, transformants with a single copy of the transforming DNA are selected. Cells transformed with this DNA construct are referred to as "substrate-transformed" cells.

Step 2. Homozygous transgenic plants containing the transforming DNA are regenerated from the selected substrate-transformed cells. These plants are

crossed with a line that expresses the transposase specific for the DNA substrates engineered into the vector (if created recombinantly, then referred to as a "transposase-transformed" line), to produce heterozygous F1 progeny. The progeny contain the transposase and the transforming DNA construct harboring the transposase recognition sites. Excision and integration events occur naturally in these hybrid plants as they grow, due to the presence of the transposase and its substrate. Since the F1 hybrids are heterozygous, excision at the DNA substrate sites on the construct occurs which releases a portion of the transforming DNA, referred to as the "recombination substrate". In a preferred embodiment the excision will generate, per cell, a single copy of the "recombination substrate". The recombination substrate contains the targeting segment with the positive selection gene, as well as the negative selection gene located outside the targeting segment.

Step 3. The F1 plants are allowed to self-pollinate to produce F2 seed. Optionally, this step may be carried forward into the F3 and subsequent generations. Seeds are germinated on growth media containing the positive and negative selection agents. Random integration of the recombination substrate into the genome results in plants that are sensitive to the negative selection agent and resistant to the positive selection agent. However, integration of the excised insert by double crossover events at the targeted locus results in plants that are resistant both to the negative selection agent and to the positive selection agent. These plants may be selected and/or maintained by their ability to survive in the presence of both selection agents, while plants containing random integrants cannot survive on the negative selection agent.

According to another aspect of the invention,

the aforementioned DNA constructs and methods may be adapted to perform activation tagging of a plant genome to create variants displaying a desired phenotype. In this case, the selection step for the progeny omits the negative selection. Instead, the plants are screened for the phenotype desired to be identified by the activation tagging method.

According to another aspect of the invention, kits are provided to facilitate performance of the targeted gene insertion or activation tagging methods described above. The kits provide one or more of the DNA constructs of the invention, along with instructions for their use in performing the methods.

Other features and advantages of the present invention will be better understood by reference to the drawings, detailed description and examples that follow.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Generalized *Agrobacterium*-based gene targeting construct. RB, Right border; LB, Left border; PLS1, polylinker sequence 1; PLS2, polylinker sequence 2; *CodA*, cytosine deaminase-encoding sequence; *Bar*, phosphinothricin acetyltransferase-encoding sequence; *GUS*, β-glucuronidase-encoding sequence; 35S, CaMV 35S promoter; nos, 3' polyadenylation sequence from the nopaline synthase gene; *Ds*, excision target (DNA substrate) for the *Ac*-dependent transposase.

Figure 2. Strategy for *Ac*-dependent production of recombination substrates *in planta* for gene targeting. The Arabidopsis gene *TGA3* is used as an example to illustrate the use of *Ac*-dependent excision to generate substrates for recombination. The hatched regions designate genome sequences flanking *TGA3* and are cloned

into the polylinker sites (PLS1, PLS2) of the targeting construct as shown in Figure 1. Homozygous transgenic plants containing this insert are crossed with a line that is expressing the *Ac* transposase (35S:*Ac*/*NptII*). Excision at the *Ds* sites of the transgene releases a single molecule of the recombination substrate. Random reintegration of the excised insert produces plants that are 5-Fc^s, PPT^r, and GUS$^-$ while the parental line is 5-Fc^s, PPT^r, and GUS$^+$. However, integration of the excised insert via double cross-over events at the *TGA3* locus will result in plants that are 5-Fc^r, PPT^r, and GUS$^-$. These can be confirmed by performing PCR with genomic DNA using the diagnostic primers P1 and P2, as indicated. The heterozygous nature of the primary targeted transformant is illustrated on the bottom by showing the wild-type *TGA3* allele along with the targeted allele.

DETAILED DESCRIPTION OF THE INVENTION

I. <u>Definitions</u>

Various terms relating to the biological molecules and other aspects of the present invention are used throughout the specification and claims.

With reference to the mutations of the invention, the term "null mutant" is used to designate an alteration in the genomic DNA sequence of an organism that can cause the product or function of the gene to be largely absent or nonfunctional. Such alterations may occur in coding and/or noncoding regions of the gene, including regulatory regions or other regions which when altered cause said product or function to be largely absent or nonfunctional. The alterations may include insertions and/or deletions of one or more base pairs and/or changes in one or more base pairs.

In reference to the strategic placement of heterologous DNA segments within the genomic DNA, the

term "targeted gene insertion" is used to designate the designed, engineered, creative and/or logical selection of specific genomic DNA sequence(s) of interest for insertion, deletion or substitution of one or more base pairs of DNA. This DNA may encode a detectable or selectable gene product or function to facilitate identifying and/or isolating successful "targeted gene insertion" events. "Targeting" is accomplished by placing "targeting DNA sequences" having homology with the known, determined or predicted DNA sequence(s) of the genomic DNA of interest into the constructs of the invention in a manner such that homologous recombination may occur. "Targeted gene insertion" may typically create a null mutant, but may also create an up-regulated or down-regulated gene, or may have no ascertainable effect on the genomic DNA so altered. The "targeted DNA" or "targeted genomic DNA" is the genomic sequence of interest from the organism to be transformed.

In reference to placement of exogenous DNA within the genomic DNA of an organism in locations other than those determined by strategic placement or designed, engineered, creative or logical selection, the terms "random insertion" and/or "random integration" are used. This DNA also may encode a detectable or selectable gene product or function to facilitate identifying and/or isolating "random insertion" events. "Random insertion" may occur by homologous or heterologous recombination with genomic DNA sequences. "Random insertion" may create a null mutant, an up-regulated or down-regulated gene, or may have no ascertainable effect on the genomic DNA so altered.

With reference to nucleic acid molecules, the term "isolated nucleic acid" is sometimes used. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately

contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a procaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule.

The terms "recombinant substrate" or "recombination substrates" refer to the DNA molecules which are produced in F1 progeny produced by the method of this invention. The recombination substrate contains the targeting sequence with the positive selection gene, as well as the negative selection gene located outside the targeting sequence. The recombinant substrates result from the excision by the transposase activity which specifically recognizes the DNA substrates in the DNA constructs of this invention. The recombinant substrates, when integrated by homologous recombination in the F2 progeny, result in organisms which have an insertion in the targeted gene and which are selected by their resistance to both the positive and negative selection agents.

The terms "DNA substrate" and "excision site" are use in reference to the specific sequences or locations within the DNA molecules at which the transposase enzyme activity of a transposable element system can excise flanking DNA sequences. These DNA substrates are also referred to herein as "transposase recognition sites". Each transposase activity has specificity for its own DNA substrate sequence in a manner that is integral to that particular transposable element system.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid,

oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest (e.g. chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

Nucleic acid sequences and amino acid sequences can be compared using computer programs that align the similar sequences of the nucleic or amino acids thus define the differences. In preferred methodologies, the BLAST programs (NCBI) and parameters used therein are employed, and the DNAstar system (Madison, WI) is used to align sequence fragments of genomic DNA sequences. However, equivalent alignments and similarity/identity assessments can be obtained through the use of any standard alignment software. For instance, the GCG Wisconsin Package version 9.1, available from the Genetics Computer Group in Madison, Wisconsin, and the default parameters used (gap creation penalty=12, gap extension penalty=4) by that program may also be used to compare sequence identity and similarity.

The terms "percent identical" and "percent similar" are also used herein. When referring to nucleic acid molecules, "percent identical" refers to the percent of the nucleotides of the subject nucleic acid sequence that have been matched to identical nucleotides by a sequence analysis program.

With respect to single-stranded nucleic acid molecules, the term "specifically hybridizing" refers to the association between two single-stranded nucleic acid molecules of sufficiently complementary sequence to permit such hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term

refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence.

A "coding sequence" or "coding region" refers to a nucleic acid molecule having sequence information necessary to produce a gene product, when the sequence is expressed.

The term "operably linked" or "operably inserted" means that the regulatory sequences necessary for expression of the coding sequence are placed in a nucleic acid molecule in the appropriate positions relative to the coding sequence so as to enable expression of the coding sequence. This same definition is sometimes applied to the arrangement other transcription control elements (e.g. enhancers) in an expression vector.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell.

The terms "promoter", "promoter region" or "promoter sequence" refer generally to transcriptional regulatory regions of a gene, which may be found at the 5' or 3' side of the coding region, or within the coding region, or within introns. Typically, a promoter is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. The typical 5' promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or

elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

A "vector" is a replicon, such as plasmid, phage, cosmid, or virus to which another nucleic acid segment may be operably inserted so as to bring about the replication or expression of the segment.

The term "nucleic acid construct" or "DNA construct" is sometimes used to refer to a coding sequence or sequences operably linked to appropriate regulatory sequences and inserted into a vector for transforming a cell. This term may be used interchangeably with the term "transforming DNA". Such a nucleic acid construct may contain a coding sequence for a gene product of interest, along with a selectable marker gene and/or a reporter gene.

The terms "selectable marker gene" or "selection marker gene" refer to a gene encoding a product that, when expressed, confers a selectable phenotype on a transformed cell. "Positive selection marker" refers to a gene whose functioning gene product, when expressed, confers upon a cell the phenotype of survival or growth in the presence of a positive selection agent which is deleterious or lethal to cells which do not possess the "positive selection marker". "Negative selection marker" refers to a gene whose functioning gene product, when expressed, confers upon a cell the phenotype of susceptibility to the presence of a negative selection agent to which cells which do not possess the "negative selection marker" are not susceptible.

The term "reporter gene" or "detectable marker

gene" refers to a gene that encodes a product which is easily detectable by standard methods, either directly or indirectly.

A "heterologous" region of a nucleic acid construct is an identifiable segment (or segments) of the nucleic acid molecule within a larger molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. In another example, a heterologous region is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains introns, or synthetic sequences having codons different than the native gene). Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein. The term "DNA construct", as defined above, is also used to refer to a heterologous region, particularly one constructed for use in transformation of a cell.

A cell has been "transformed" or "transfected" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A

"clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

Other definitions may be found in the description that follows.

II. Description

To practice the novel gene targeting strategy of the present invention, a DNA construct for transforming cells is needed that has, combined in novel fashion, the following elements: (1) targeting segments that comprise extended regions of homology with the targeted location on the genome; (2) a positive selection gene contained between appropriate targeting segments; (3) a negative selection gene located outside the targeting segments; and (4) a pair of DNA substrates for a selected transposase, located outside the targeting segments and the negative selection gene. Optionally, the region between the targeting segments may also contain one or more cloning sites for insertion of additional nucleotide sequences. In addition, the transforming DNA construct optionally may contain a reporter gene which, if present, preferably has disposed therewithin one of the transposase recognition sites, such that upon excision, the activity of the reporter gene product is not detectable. In a preferred embodiment, the DNA substrate is located between the promoter sequence and the gene sequence encoding the reporter gene product, such that upon excision and reintegration, the intact promoter remains nears one of the ends of the integrated DNA. In another preferred embodiment, the DNA substrate is short (<1.5 Kb) such that it still retains the specific recognition sites for the transposase, but does not interfere with the ability

of the promoter to drive expression of the reporter gene activity.

The targeting segment is a DNA sequence that has homology with a selected region of the genome being transformed, and which is of sufficient length and homology to ensure the homologous recombination event necessary for incorporation of the transforming DNA into the genome. The targeting segment may comprise regions of homology that encompass or flank the selected target region of the genome. The targeting segments are preferably greater than 500 bases on either side of the positive selection gene and optional additional nucleotide segments. More preferably, they are greater than about 1 kb to 1.5 kb on either side of the positive selection gene, and most preferably they are at least 2-3 kb on either side of the positive selection gene. One skilled in the art will be able to determine the required length of the targeting sequences by considering the relative factors of length and relative homology with the known or anticipated sequence of the targeted genomic sequence, the critical factor being that the selected targeting sequence allow for the low frequency event of homologous recombination. For example, it would be appreciated by one skilled in the art that the targeting sequence could be shorter in cases where there is a high degree of similarity or identity with the targeted genomic sequence, or that the targeting sequence might be longer in the case of low similarity or where the sequence of targeted genomic DNA is not fully known.

The targeting segments can be selected for homologous recombination with any portion of a genome of interest. Preferably, however, genome targets comprising genes or regulatory regions are selected. Alternatively, regions adjacent to or near genes may be selected, such that insertions may be made without disrupting gene

expression.

The positive selection gene may comprise one of many such genes known and used in the art. Useful selectable marker systems include, but are not limited to: genes that confer antibiotic resistances (e.g., resistance to kanamycin, hygromycin or bialaphos) or herbicide resistance (e.g., resistance to sulfonylurea, phosphinothricin, or glyphosate). In the preferred embodiment taught in Example 1, the *Bar* gene, which confers resistance to herbicides that are based on phosphinothricin (PPT), is utilized.

The negative selection gene also may be one of several such genes known in the art. Preferred for use in the invention is the *CodA* gene, encoding cytosine deaminase. This enzyme converts the innocuous 5-fluorocytosine to the cytotoxic 5-fluorouracil. Other negative selections that can be used in the invention include, but are not limited to, the *aux-2* gene from the Ti-plasmid of *Agrobacterium*, the *TK* gene from SV40, cytochrome f450 from *Streptomyces griseolus*, the *Adh* gene from maize or *Arabidopsis*, or any gene encoding an enzyme capable of converting innocuous substances into harmful or lethal substances.

The strategies of the present invention can be used in any system known now or discovered in the future to harbor transposable element systems. The present invention exemplifies gene targeting in plants using the well-characterized *Ac/Ds* system. Other plant transposable element systems suitable for use in the present invention include, but are not limited to: *Spm(En)/dSpm* from maize, *Dt/rdt* from maize, *Mu-MI/Mn* from maize, and *Tam1/Tam2* or *Tam3/Tam4* from snapdragon. The *P* element from *Drosophila melanogaster* is also suitable for use in the present invention. Persons skilled in the art

will appreciate that numerous other organisms, including *Drosophila*, yeast, the nematode, *C. elegans*, and mammals such as mice, contain characterized transposable element systems, each of which has potential for use in the present invention.

The optional detectable marker gene may be selected from any of the numerous genes known and used in the art for this purpose. Examples of suitable detectable marker genes include, but are not limited to, genes encoding: β-glucuronidase (GUS), β-galactosidase, chloramphenicol acetyl transferase (CAT), various transcription factors, alcohol dehydrogenase and luciferase. In the preferred embodiment taught in Example 1, the GUS marker is utilized.

The above-described DNA constructs may be used directly or as part of a vector, in accordance with the wide variety of transformation methods available to persons of skill in the art.

In one preferred embodiment, the gene targeting strategy of the invention is applied to plants. Transgenic plants can be generated using standard plant transformation methods known to those skilled in the art. These include, but are not limited to, biolistic DNA delivery (i.e., particle bombardment), *Agrobacterium* vectors, PEG treatment of protoplasts, UV laser microbeam, gemini virus vectors, calcium phosphate treatment of protoplasts, electroporation of isolated protoplasts, agitation of cell suspensions with microbeads coated with the transforming DNA, direct DNA uptake, liposome-mediated DNA uptake, and the like. Such methods have been published in the art. See, e.g., Methods for Plant Molecular Biology (Weissbach & Weissbach, eds., 1988); Methods in Plant Molecular Biology (Schuler & Zielinski, eds., 1989); Plant Molecular Biology Manual (Gelvin, Schilperoort, Verma,

eds., 1993); and Methods in Plant Molecular Biology - A Laboratory Manual (Maliga, Klessig, Cashmore, Gruissem & Varner, eds., 1994).

The method of transformation depends upon the plant to be transformed. The biolistic DNA delivery method is useful for nuclear transformation of monocotyledenous plants, such as maize. Alternatively, *Agrobacterium* vectors, particularly superbinary vectors such as described by Ishida et al. (Nature Biotechnology 14:745-750, 1996) are used for transformation of plant nuclei.

In another embodiment, the DNA constructs of the invention are used for activation tagging of plants. After transformation using standard plant transformation methods known to those skilled in the art, selection with the positive selection agent, but not the negative selective agent will allow the higher frequency random integration events to be recovered as well as other transformants. These transformants will also have detectable amounts of reporter gene product activity. Transgenic plants are then regenerated from these transformants and these plants are then crossed with lines expressing the transposase activity corresponding to the DNA substrates of the transforming DNA vectors. The active transposition of the integrated DNA will result in progeny with insertions in many different locations throughout the genomic DNA. These progeny will often contain genes which are being overexpressed due to transactivation by a promoter in the integrated DNA which transactivates downstream gene expression. This population of transformed and transposed mutants or its progeny can then be screened for useful phenotypes. For plants of agronomic interest, this method of activation tagging has substantial advantages over current methods in that only small numbers of independent transformants

are required, yet the novel application of the transposable element system enables the possibility of a large population of potential mutations. In a preferred embodiment of this method, the DNA substrate of the DNA construct is inserted in the reporter gene between the promoter sequence and the encoding sequence such that upon excision, the promoter will be close to one end of the excised and reintegrated DNA, and the detectable activity of the reporter gene will be largely absent or nondetectable. In a still more preferred embodiment of this method, the DNA substrate used is a short (<1.5 kb) Ds element from maize. Without intending to limit the invention in any way by explanation, presumably longer Ds elements contain transcription termination signals which would interfere with the expression of both the reporter gene as well as any transactivated genes. In a highly preferred embodiment, the multiple copies of a promoter are used or promoter(s) with inducible activity or tissue-specific activity or other such promoters as would be known to one skilled in the art to be useful. In another preferred embodiment, seed can be collected from the transformed and transposed population or its progeny to be used for screening for useful phenotypes.

The following example is provided to describe the invention in greater detail. It is intended to illustrate, not to limit, the invention.

EXAMPLE 1
Transposon-Based Gene Targeting Strategy
for Plants, Using *Agrobacterium* Vectors

This example describes new DNA targeting constructs to facilitate the transfer of gene targeting technology from *Arabidopsis* to crop plants such as maize. The new construct design comprises a more general

positive selection marker than used in current systems, as well as a substrate-dependent negative selection marker to streamline the detection of the desired event. The major goal in this new gene targeting strategy is to avoid the need to directly generate large numbers of independent transformation events via *Agrobacterium tumefaciens*. An alternative strategy utilizing the maize *Ac/Ds* transposon system is therefore employed to ascertain its efficacy of generating the substrates for homologous recombination *in planta*.

Objectives

1) General positive selection marker for plant transformation. In the inventor's earlier generation of gene targeting construct, the *NptII* gene was used as the positive selection marker, providing resistance to the antibiotic, kanamycin. Although this works well in *Arabidopsis* and many dicotyledenous plants, it is not efficient for selection in many monocotyledenous plants that have been tested. The *Bar* gene, which confers resistance to herbicides that are based on phosphinothricin (PPT), has been tested as an alternative marker. In addition to being an established positive selection marker for dicots as well as monocots, the *Bar* gene can also provide selection on soil-grown plants by herbicide spraying. Thus, it is a more versatile positive selection marker. This example describes the construction of a new generation of gene targeting vectors based on the *Bar* gene.

2) Incorporation of negative selection markers. In order to streamline the efforts in identifying the desired recombination event, a substrate-dependent negative selection marker, cytosine deaminase (*CodA*), is used. When driven by the CaMV 35S promoter, the *CodA*

gene provides negative selection during seed germination and early seedling growth in the presence of 5-fluorocytosine (5-Fc) [11, 13]. This construct was tested in *Arabidopsis* and was found to provide good negative selection on agar plates supplemented with 5-Fc (Figure 4), as reported earlier by other researchers. Incorporation of the *CodA* expression cassette into the targeting construct is intended to help to minimize the number of random insertion events.

3) Application of a transposon-based gene targeting strategy. Although, in principle, the inclusion of a negative selection marker should simplify the gene targeting approach, it may not be compatible with the preferred *Agrobacterium*-mediated transformation method. With *Agrobacterium*-based plant transformation strategies, multiple copies of the T-DNA are often inserted in the genome of transformed plant cells. In this case, the targeted event may coexist with random insertions and is removed when negative selection is applied (5, 13).

To avoid this problem, one solution is to devise a method to limit the substrate for recombination to one copy per cell. This should then rule out the possibility of having multiple insertion events, through either illegitimate or homologous recombination. To accomplish this, the *Ac/Ds* transposon system is employed as an *in planta* generator of integration substrates (12). This is described in more detail in the following section.

In addition to solving the problem of mutiple integration events, this approach also broadens the application of gene targeting to other plant species. A major obstacle in applying gene targeting to agronomically important species, such as maize and rice,

is the difficulty in generating a large number of independent transformants (in the order of 1,000). Novel use of a transposon-based gene targeting method eliminates this difficulty.

Experimental Approach

The experimental design utilizes a known genomic target in *Arabidopsis*. The general targeting construct shown in Figure 1 is constructed, using standard cloning and DNA manipulation methods. As shown in Figure 1, the *CodA* gene is placed upstream of the polylinker region, into which can be inserted the genomic sequence for targeting. For DNA excision, short (<1.5 kb) Ds elements are placed next to the Right Border and within the 35S-GUS cassette. With this configuration of the construct, the frequency of excision can be assayed by measuring the loss of GUS activity, since this loss results in white sectors upon staining with X-Gluc. The *Bar* gene, flanked by the inserted genomic sequences, provides the positive selection marker for the insertion event. After construction of this vector, *Arabidopsis* genomic sequences from *TGA3* are inserted into the polylinker sites (Fig 1, PLS1 and PLS2). Transformation of *Arabidopsis* via *Agrobacterium* is carried out and the transgenic lines are selected on PPT-containing plates.

These PPT-resistant plants, but not the wild-type, are also sensitive to 5-Fc, due to *CodA* expression. In addition, they have high levels of GUS activity (Figure 2, step 1). Ten to twenty transformed lines that show these characteristics are evaluated by Southern blot analysis to determine the copy number of the inserted T-DNA. Several transformed lines with single copy insertions are self-pollinated to produce homozygous plants. They are then crossed with another homozygous

Arabidopsis line that expresses the stablized *Ac* transposase (12). This activates the excision of the sequences in the original construct that are flanked by the two *Ds* elements (Figure 2, step 2). The efficiency of excision in the F1 progenies is verified by staining the leaves with X-Gluc. Cells that have activated *Ds* transposition are white while the other cells are blue. The excised DNA can be reinserted randomly in the genome, in which case the *CodA* gene is retained (Figure 2, excised insert). Alternatively, if the insertion occurs via homologous recombination at the two homology regions, then this marker is lost. The selection scheme entails screening the F2 or F3 progenies on medium containing 5-Fc and PPT. The surviving plants are then subjected to a secondary screen to confirm the loss of the GUS marker. The predicted genome structure and phenotypes of the targeted event are shown at the bottom of Figure 2. Plants recovered after these screens are analyzed via PCR using diagnostic primers (shown as P1 and P2 in Figure 2) that specifically detect the desired targeted event (7). Southern blots are performed to confirm the proper integration of the *Bar* gene into the *TGA3* locus.

REFERENCES

1) Bensen RJ, Johal GS, Crane VC, Tossberg JT, Schnable PS, Meeley RB and Briggs SP (1995) **Plant Cell** 7: 75-84.

2) Capecchi MR (1989) **Science** 244: 1288-1292.

3) Kempin SA, Liljegren SJ, Block LM, Rounsley SD, Yanofsky MF and Lam E (1997) **Nature** 389: 802-803.

4) Krysan PJ, Young JC, Tax F and Sussman MR (1996) **Proc. Natl. Acad. Sci. USA** 93: 8145- 8150.

5) Liljegren SJ and Yanofsky MF (1998) **Trends in Plant Sci.** 3: 79-80.

6) McKinney EC, Ali N, Traut A, Feldmann KA, Belostotsky DA, McDowell JM and Meagher RB (1995) **Plant J.** 8: 613-622.

7) Miao Z-H and Lam E (1985) **Plant J.** 7: 359-365.

8) Paszkowski J, Baur M, Bogucki A and Potrykus I (1988) **EMBO J.** 7: 4021-4026.

9) Puchta H (1998) **Trends in Plant Sci.** 3: 77-78.

10) Risseeuw E, Franke-van Dijk MEI and Hooykaas PJJ (1997) **Plant J.** 11: 717-728.

11) Stougaard J (1993) **Plant J.** 3: 755-761.

12) Sundaresan V. et al. (1995) **Genes and Dev.** 9: 1797-1810.

13) Thykjaer T, Finnemann J, Schauser L, Christensen L, Poulsen C and Stougaard J (1997) **Plant Mol. Biol.** 35: 523-530.

14) Walden R, Fritze K, Hayashi H, Miklashevichs E, Harling H and Schell J (1994) **Plant Mol. Biol.** 26: 1521-1528.

15) Matsuhara S, Jingu F, Takahashi T, Komeda Y (2000) **Plant J.** 22: 79-86.

The present invention is not limited to the embodiments described and exemplified above, but is capable of variation and modification without departure from the scope of the appended claims.